# Knowledge Discovery and Dissemination (KDD) BAA Questions

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 001 | Is it anticipated that the number of new and/or surprise data sets will be small enough to permit manual involvement of contractor personnel sufficient to complete the 'n-choose-2' alignment activities within the time allotted for 'set up' during an evaluation? | No. There will be a time limit and a limit on the number of personnel that can be used during the alignment phase of the evaluation. This will effectively prevent the alignment from being done completely manually. See Section 1.A.4.1 (The KDD Cycle). | 1/15/10 |
| 002 | In reference to Section 1.B (Data Description), could a data set also consist of a collection of unstructured natural language English documents, whose understanding will require the performer to do natural language processing? | Data sets can be structured or unstructured; data sets may be a product of transcription or a translation. See Section 1.B (Data Description) and also Section 1.A.3 (Description of Desired Research/Research Not of Interest). | 1/15/10 |
| 003 | Can a research group in a French University take part in the IARPA KDD program? | See Section 3.A (Eligible Applicants) and 3.D.1 (Collaboration Efforts). | 1/15/10 |
| 004 | Must all software be developed in Java? Is the source code required to be delivered? | The BAA has been amended to allow software for alignment to be free of the Java-only restriction. For the advanced analytics, all new software must be written in Java, but it is permitted to use the Java Native Interface (JNI) ("wrappers") to utilize legacy software that may have been written in other language(s). As stated in Section 4.B.3/Section 3 (Detailed Proposal Information), "all source code and the appropriate scripting, subordinate libraries, release notes, and other necessary components, data and documentation, must be delivered." | 1/26/10 |
| 005 | We operate under a Special Security Agreement, which was executed by the Defense Security Service as the | A determination will be made based on the information submitted in both the Appendix F - SF 328 (Certificate | 1/29/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | foreign ownership, control, operation, or influence (FOCI) mechanism. Do we meet the provisions stated on page 32 of the Solicitation with regard to FOCI? | Pertaining to Foreign Interest) and Appendix G-KMPL (Key Management and Personnel Listing). | |
| 006 | Does IARPA anticipate this award to include any implementation requirements, or strictly R&D? | The awards will be for research and development activities. | 1/29/10 |
| 007 | Should the Section 3.A.1 OCI waivers or certifications be included as part of the Cost Proposal or Technical Proposal? | See Section 4.B.1 - Volume 1: Technical and Management Proposal. | 1/29/10 |
| 008 | In reference to Section 1.B (Data Description), can you provide guidance on the likely size of unstructured text data sources? | Unstructured data sets have the same general guidance as structured and will be in the range of up to roughly 10,000,000 records. In context of unstructured data, examples of "records" could be articles, newswires, messages, or words. | 1/29/10 |
| 009 | In reference to Section 4.B.1/Section 4 (Security Plan) instructions for submission of foreign ownership, control or influence (FOCI) via the SF328 form, if the form is filed with the Department of Defense, do you need all the schedules that go along with the FOCI filing? | Yes. | 1/29/10 |
| 010 | By not modifying "the content of any provided data set", is it correct to believe that an integrity-verifying hash function of an original data file must validate on the same data file after alignment is complete? Extracting information from a data set in a read only mode and creating further refinements of it in a separate file for alignment purposes does not violate this restriction? | Yes to the first question and correct to the second question. | 1/29/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 011 | If all the data files within a data set are aligned by creating (possibly large) RDF files, what are the characteristics of the triple store that will hold these RDF statements for later exploitation? Is this determined by the test configuration, or will performers be allowed to bring triple stores to the exercise? If the former, will it have RDF and OWL-DL inference rules? Will it be capable of accepting additional inference plug ins? | The characteristics will be either: a) MYSQL based SDB Jena store or b) HADOOP HTFS with HBase

The performers will not be allowed to bring their own versions of triple store. In the first year, we are trying to get algorithms to work on a common platform rather than prematurely optimizing triple store performance and capacity. OWL-Full allows for inclusion of inferences. Any inference plug-in must be compatible with the Jena (and BLACKBOOK) reasoning API. | 1/29/10 |
| 012 | In preparing an OPSEC section to the Security Plan, can we receive a copy of the Critical Information List - and if so, can we receive a point-of-contact? | No, a copy of the Critical Information List will not be received until after award. See Section 4.B.1/Section 4 (Security Plan). | 1/29/10 |
| 013 | Is the following within scope of the BAA: The development of technologies that permit analysis across multiple data sets that cannot be simultaneously analyzed for policy reasons (e.g. by anonymizing results of intermediate analyses)? | Development of such technologies is not within scope. | 1/29/10 |
| 014 | How do we confirm accreditation of Information Systems if we do not have systems approved for the KDD Program and would not at this stage of the contract? Do you want us to show that we have other accredited Information Systems for other Government classified programs? | As outlined in the BAA, at the time of proposal, each offeror is required to provide certification that they have an accredited SECRET facility clearance for all facilities in which they intend to process classified information. In addition, if an offeror chooses to use an Automated Information System (AIS) for the alignment prototype in the KDD evaluation that is different from the GFE specified in Appendix E, the offeror must provide certification that the system intended for use has been accredited for the | 1/29/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | | processing of information at the SECRET level, and is suitable for the processing of SECRET//NOFORN information. These certifications, both of the proposed facility (ies) and of the AIS, will be used by the KDD Program to secure the necessary co-use agreements required to support this program. | |
| 015 | The BAA says that "The amount of resources made available under this BAA will depend on the quality of the proposals received and the availability of funds." Can you expand on the scope envisioned for the awards? | No. | 2/4/10 |
| 016 | Will all teams get the same challenge problems / datasets for all four evaluations? | Yes. | 2/4/10 |
| 017 | In reference to Section 1.B, will the real and surrogate datasets be matched in size and complexity? | The real and surrogate datasets will be matched in size and complexity to the extent possible. | 2/4/10 |
| 018 | Is it perceived to be an advantage if a team has previous experience with BLACKBOOK? | BLACKBOOK is a central component in KDD's evaluation plan. BLACKBOOK software and documentation will be provided to all performers, along with a limited amount of support. BLACKBOOK experience is clearly useful, but is not required. | 2/4/10 |
| 019 | In reference to Section 1.A.5.1, does the performer alone need to have OWL / Semantic Web expertise since they will work most closely with the evaluation process? | The performer team needs appropriate OWL/semantic web expertise. | 2/4/10 |
| 020 | Does the proposer need to construct an ontology representing the high level semantics that informs the explicit or implicit data models for the datasets? | This is part of the alignment process and is left for the performer to develop a solution. | 2/4/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 021 | If a proposer identifies a new relevant dataset, or newly acquires access to one, during the course of the program, will there be any chance for allowing the team to use this dataset for the proposed work? | The performer will need to provide written verification as instructed in Section 1.B (Data Description), prior to use of any new data in pursuit of their research. | 2/4/10 |
| 022 | Section 1.A.3 enumerates 10 research areas not of interest to KDD, including "10) Research on Resource Description Framework (RDF) extraction." Can you clarify what research is ruled out here? | The work of RDF extraction is part of the task; however, it is not a major research focus of KDD. KDD research may include creating RDF models between different schemas. | 2/4/10 |
| 023 | How much structure within an input document is required for ontology/data model extraction to be successful? Does the data have to be in a structured form, like a spreadsheet, or can it exist in a more unstructured form, like raw text? Do we have to alter the ontology extraction algorithm in accordance with the structure of the document? | This is part of the alignment process and is left for the performer to develop a solution. | 2/4/10 |
| 024 | What exactly are we extracting from an input document? Are we extracting entire ontologies, instances, individual concepts, or some combination of the above? | This is part of the alignment process and is left for the performer to develop a solution. | 2/4/10 |
| 025 | Are we going to know the source of the input documents? To what degree would this affect the ontology/data model extracted? | Performers will be provided with information about the source of the input documents to the extent possible. How the performer uses this is part of the alignment process and is left for the performer to develop a solution. | 2/4/10 |
| 026 | How do we know when to derive a new ontology, or when to fit the document to a preexisting ontology? Should a preexisting ontology be changed to accommodate a document that almost fits into it? If so, then when exactly should this occur? | This is part of the alignment process and is left for the performer to develop a solution. | 2/4/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 027 | For ontology alignment, can we assume that we will always start with a set of input documents that are populated with data as opposed to having only a single input document example with data or a single example without specific data? | The data sets provided can be populated excel spreadsheets, messages, text documents and common separated values (csv), etc. The data sets will all be populated. | 2/4/10 |
| 028 | Will any government supplied target ontologies be limited to a set of OWL 2 profiles, or should be prepared to handle any OWL profile? | The KDD Program will not provide a target ontology. | 2/4/10 |
| 029 | In terms of knowledge domains, what level (granularity) of knowledge are we going to consider? For instance, would we limit ourselves to the domain "sports," or would we consider more specific domains, such as "baseball," "basketball," etc? | There will be domain limits. Any domain limits will be defined implicitly by the challenge problem and the data sets provided | 2/4/10 |
| 030 | Are query building or question answering interfaces of interest? | See Section 1.A. 3, Item 9. This work is not a research interest of KDD; however, the offeror may have technologies in this area that would be useful, and is not prohibited from proposing to use them. | 2/4/10 |
| 031 | In reference to Section 4.A.1, the BAA states that the Government will use several external companies to assist in evaluation of the proposals. If our cost proposal is to be viewed by any entity other than the Government, please provide contact information for each of those other entities so we may sign NDA's with them. | See Section 4.A.1 (Proposal Information) for "notice of objection" instructions. | 2/4/10 |
| 032 | The Base of 15 months will be bid as a Firm proposal but for the 3 Options are they allowed to be bid as ROM's or must they be Firm's? | The base and the three options must all have Firm cost proposals against them. | 2/4/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 033 | If we are required to use GFE (Government Furnished Equipment) or GFP (Government Furnished Property), are we expected to obtain estimates for these costs and include them in our cost proposals? Are the GFE or GFP costs considered part of the total contract award? | No. GFE/GFP will be provided to the awarded-contractors at no cost. | 2/4/10 |
| 034 | The BAA does not specify when GFE will be delivered to the performers. Can you please clarify when this is expected to take place; e.g. in month 1, or in month 9? | GFE will be delivered within one month of award. | 2/4/10 |
| 035 | Does the GFE come with a removable hard drive? | Yes. | 2/4/10 |
| 036 | Would we need to price the cost for shipping the GFE back? Or would the government cover that cost? | You do not need to price the cost for return shipment. | 2/12/10 |
| 037 | Could you please advise roughly how much time each performer will get in the government's test facility? Roughly how much of the evaluation week will be allowed for alignment in the Pre-Test evaluation? | For the Pre-Test Evaluation cycle, each performer should plan for up to a total of three (3) weeks in the Government's test facility. Within this 3 week period, performers can dedicate up to two (2) weeks for alignment and test preparation, and can then expect one (1) week for actual evaluation. For subsequent evaluation cycles, the performers should expect the time needed for alignment and test preparation to be reduced to a single week. | 2/12/10 |
| 038 | In reference to Appendix D, if the contractors have existing analysis algorithms encoded in other languages, can they interface them to BLACKBOOK via web services or other mechanisms rather than spending the time to recode their algorithms? If complete translation to Java is desired, could the translation be postponed to later in the program, after the best approaches have been selected? Does this prohibit the use of existing commercial or proprietary code, | See Question 4 and BAA 09-10, Amendment 1. See also Section 6.B.3.1.2 (Commercial Items) for commercial products (e.g. MATLAB, Microsoft), the performer must identify the product and the cost per license. See Section 6.B.3.1 (Noncommercial Items) for non-commercial software (e.g. proprietary software). | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|---|---|---|
| | where source code is not available for delivery, in the solution? What if the executable of the proprietary code is available free for Government use? | | |
| 039 | Are the 10 questions on page 6 examples of what needs to be encoded into queries or samples of natural language queries that must be processed? | These are examples used solely to illustrate the classes of query. | 2/12/10 |
| 040 | Question 5 on page 6 of the BAA asks for what is near a particular location. Should we assume that the correct answer will be in the supplied data or do we need to provide the application with its own geographical knowledge base? Given that the GFE only comes with 2 Gbytes of disk space and there are no limits on the query, there may not be sufficient disk space. | You can assume that information to answer questions will be in the data provided. When the KDD Program identifies the challenge problem for the evaluation cycle, performers can request permission to use specific supporting data. Also see Appendix E, the GFE hardware includes at least 4TB of hard drive space. | 2/12/10 |
| 041 | Roughly how large could the test data sources be? How much GFE hard disk space can we rely on having after all the KDD data, software, and logs are counted? | See Section 1.B (Data Description) and Appendix E for the GFE description. | 2/12/10 |
| 042 | Should we assume that there will be retractions of the facts provided in the source data? For example, will the analyst decide that one of the original statements is wrong and delete it from the RDF graph? | This may be part of the alignment and advanced analysis processes. However, the original data may not be modified. | 2/12/10 |
| 043 | We are not allowed to change the original data although we are allowed to reformat it into RDF. Can we change the labels/classes on the data as part of the reformatting or should there be a one to one correspondence between, say, the column names of an RDB table and the RDF properties? | As part of the alignment process, labels and classes are mapped to one another; for example, city may be mapped to town. The details of this process are left to the performer to develop a solution. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 044 | What assumptions can we make about the formats of the data sources? | You cannot assume particular formats. At the beginning of each cycle, the format of all the relevant data sets will be provided. The performers may expect formats such as SQL/ODBC, Microsoft Excel, and comma separated values, but there may be exceptions. | 2/12/10 |
| 045 | On page 6, does "understanding of what the analyst would view as interesting" refer to the particular user doing the evaluation or a typical user? That is, does the system need to build a model of the analyst using it? | Given the high level analysis problem and specific analytic tasks, it should be possible to develop what information would be useful to a typical user. We are not interested in modeling a particular user. | 2/12/10 |
| 046 | Will the data describe "real world" entities and actions or could it be purely abstract (symbolic)? | The data will be about real world entities and actions. | 2/12/10 |
| 047 | Page 14 specifies a limited inclusion of foreign language material. Which languages will be used? Character sets? | See Section 1.B (Data Description). All character sets will be Roman. At the beginning of each cycle, the foreign language formats will be provided, but "foreign language text will only be included as addresses, names, and locations." | 2/12/10 |
| 048 | Do the alignment and analysis prototypes need to be able to run independently or can the alignment depend on the analysis algorithms? | See Section 1.A.5.1 (Alignment Process and Alignment Prototype), Number 5. This is part of the alignment and advanced analytic process and is left for the performer to develop a solution. | 2/12/10 |
| 049 | In reference to 1.C. 3 KDD Program Timeline, will the government consider shifting the month 10 prototype delivery to month 11 in order to provide consistency with the subsequent delivery schedule? | No, the first delivery will remain at month 10. The awarded performers should plan to continue research during the intervals between submitting prototypes and beginning evaluations. | 2/12/10 |
| 050 | In reference to Appendix D (BLACKBOOK), does the bidder need to have experience with BLACKBOOK in | See Section 1.A.5.2. (Analysis Process and Advanced Analytic Prototype). The bidder does not need to have | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | order to be considered? If so, how much experience must the bidder/team possess with the framework? Can the bidder propose and develop a system that can be plug-in capable to the BLACKBOOK because of the bidder's experience in developing systems that are API capable without demonstrating a significant understanding of the BLACKBOOK framework? | experience with BLACKBOOK to bid. Advanced analysis software needs to be developed and delivered in the BLACKBOOK framework to be accepted for evaluation. This means advanced analytic algorithms must be delivered as BB compliant services and not APIs into another framework system. See also Section 1.A.4.1 (The KDD Cycle); "the performers will be provided up to 200 hours of BLACKBOOK technical assistance during the first year of the KDD Program." | |
| 051 | Section 5 defines the award instrument requested as "Cost Reimbursable contract." FAR Part 16.3 lists several types of "cost reimbursable" contracts. Are we correct in our understanding that the government wants KDD offerors to request a specific type of cost reimbursement contract for the type of work that is to be performed? | Consistent with BAA Part 1 and Part 2 (Sections 2 Award Information & Section 4.B.2 Volume 2: Cost Proposal), the Government anticipates the award of a procurement contract. Procurement contracts may include cost reimbursable type vehicles as described under FAR 16.3. | 2/12/10 |
| 052 | If the prototype transforms a data set into RDF, and includes a component which corrects, de-duplicates, or fills in missing data, does this violate the "shall not modify" clause? | No. The restriction is that the prototype cannot modify the original data sets provided. The performer may modify any created file by corrections, de-duplicates, filling in data, etc. | 2/12/10 |
| 053 | Are the transformations (e.g., into RDF) performed during the execution of the analytic algorithms only? Is it permitted to use a pre-processing step to normalize the data into RDF before ingestion by prototypes? | It is permitted to use a preprocessing step to normalize the data into RDF before ingestion by the prototypes. | 2/12/10 |
| 054 | On pg. 12, it states that the alignment process may convert file formats. Would the alignment process transform the data into RDF in addition to creating OWL-2 statements? | The performer is allowed to transform the data into RDF prior to alignment, during alignment or after alignment, and it is left to the performer to develop a solution. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 055 | If so, and the RDF produced is not valid RDF (according to the created ontology), is it permitted to modify the RDF triples generated as part of the alignment process? | Yes. See Section 1.A.5.1 (Alignment Process and Alignment Prototype), Number 2 instructions that the result of the alignment process must be a collection of OWL-2 statements. | 2/12/10 |
| 056 | In reference to Section 1.B (Data Description): What is the definition of data warehouse in the paragraph? Is this limited to SQL based RDBMS technologies? If two original data files are aligned and used to create two RDF file that use the same namespaces and refer to each other's classes, properties, and instances, is this not a "Data warehouse that merges multiple data sets together?" | By "data warehouse" we mean a system in which input data sets are merged. It is not permitted to merge data sets, using any technology. If two original data sets are aligned and used to create one RDF file for each of the two original data sets, and these two RDF files use the same namespaces and refer to each other's classes, properties, and instances, we do not consider this a data warehouse. However, the two RDF files are subject to the same restriction as the two original data sets, and must remain as separate files. | 2/12/10 |
| 057 | In reference to Volume 1/Section 4 Security Plan requirements for information, please let us know how to provide personnel information in a way that will reflect the sensitive nature of the information. Specifically, can we submit this information through a secure website or secure E-mail address? | No. The Security Plan is reviewed only by the government evaluation team and will be provided a level of protection commensurate with any proprietary claims (see Section 6.B.2), the Privacy Act, and any classification controls. The Security Plan should be submitted in the same package as the rest of the proposal. | 2/12/10 |
| 058 | For those subcontractors who are U.S.-based Institutions of Higher Education, are they also required to complete the SF 328? Does this hold for those who will be performing only unclassified work? Is there guidance on how academic institutions should complete the SF328 and fill out their KMP List, as these forms seemed tailored to firms (private businesses) that are assumed to hold facility security clearances. | See Section 4.B.1/Section 4. All offerors (primes), and any subcontractors that will work with classified information, must submit a Standard Form 328, Certificate Pertaining to Foreign Interest, and a Key Management/Personnel Listing (KMPL) with their proposal. For the subcontractors who will NOT be working with classified information, these forms do not need to be supplied, but the prime MUST document in the proposal for each such subcontractor that they will not work with classified information. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 059 | In reference to Section 4.B.1/Security Plan, what is meant by a SECRET//NOFORN level facility and Information Systems for all facilities proposed to use: | KDD BAA amendment 2 corrects the requirement for SECRET-level clearances for Facilities and Information Systems. (See 1.A.2, 3.D.1, 4.B.1, and 5.A.6.) | 2/12/10 |
| 060 | In reference to Pre-Publication Review instructions, if data sets are not classified, what are the terms and conditions of such pre-publication reviews (how many months prior notice needs to be given and are Government contractors going to review them)? What procedures and criteria apply to US academic institutions? Please specify contractual requirements related to this issue that an industrial prime contractor would have to include in a sub-contract to a US academic institutions to comply with this issue. | See Section 6.B.6 (Publication Review). All publications on research supported by the KDD Program must be provided to the KDD Program Manager for pre-publication review. Government personnel will review the articles; such reviews typically require 20 working days. The prime contractor will be required to flow-down the pre-publication review requirement to all of their subcontractors. | 2/12/10 |
| 061 | Is DOD-Secret sufficient to handle the classified component of this program? Is it expected that all classified data sets will be SECRET and NOFORN at the same time? | Offerors should assume that all classified data in the KDD program will be at the SECRET//NOFORN level. Therefore, all KDD performers receiving access to classified data must have an existing SECRET clearance and must be able to receive and handle SECRET//NOFORN materials. SECRET clearances held by DOD or other eligible government agencies are acceptable for purposes of proposal submission. By definition, NOFORN material is not releasable to Foreign Nationals, Governments, and/or Non-US Citizens. | 2/12/10 |
| 062 | Will IARPA sponsor people to get clearance for members of successful bid teams? | See Section 4.B.1/Section 4 (Security Plan). Team members who support classified work must hold appropriate clearances at the time of proposal submission and the offeror must have a sufficient number of cleared personnel at the time of proposal submission to execute the offeror's proposed plan. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 063 | If we have the SF328 form on file with the Defense Security Service (DSS) which is part of the DOD, do we still need to fill out the SF328 again and submit to IARPA? | See response # 9 and Appendix F. The SF 328 must be submitted, along with all supporting documentation. | 2/12/10 |
| 064 | Do we need to separate classified work from unclassified work in our cost estimate? | This is preferred, but not required. | 2/12/10 |
| 065 | Assuming GFE does have a removable hard drive, after the contract ends, can we treat the rest of GFE (excluding the hard drive) as 'controlled' but not classified equipment? | After all non-volatile storage, including the hard drive(s) have been removed, then this is correct. The government will determine the disposition of all GFE when a contract ends. | 2/12/10 |
| 066 | In reference to Section 1.B "The performer can bring supporting data for the alignment process of a KDD evaluation", will the KDD evaluation have access to publicly accessible websites, such as Wikipedia or Princeton Wordnet? These might be useful resources for data model alignment. | See Section 1.A.5 (Technical Description of the KDD Prototypes and Their Use): The prototypes will not have access to the Internet or to any other networks during the evaluation process. | 2/12/10 |
| 067 | Are the OWL-2 statements produced by the data alignment process, statements about the data itself (e.g., "ChevyCavalier hasSize mid-size") or is it statements about the data files and metadata (e.g., "CarModels.xsl hasFormat Excel")? I assume the former, as the latter would be rather limited. | Section 1.A.5.1 (Alignment Process and Alignment Prototype). The former is correct, although OWL/RDF can certainly represent metadata as well. | 2/12/10 |
| 068 | Not all concepts are easily represented by OWL ontology - for instance, geographic inclusion (Iraq hasBorder [lots of geo coordinates here]). Therefore I assume that not all rules and analysis used by the analytic algorithms must be specified in OWL-2. Correct? | Output from the alignment algorithms may only be expressed as OWL-2 statements. While there may be some concepts that are difficult to completely express in OWL-2, adopting OWL-2 will enable the government to mix and match tools, fully leveraging all of the funded work in the program. Offerors should keep in mind that the emphasis of | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | | KDD is on the combination of analysis with alignment; therefore, offerors should propose approaches that will produce answers to analytic questions in situations where data are not perfectly aligned. That said, the Government intends to choose data sets and challenge problems where the needed expressiveness should be within the capabilities of OWL-2. | |
| 069 | For a particular analytical challenge, will the 'ground truth' be available either before or after the evaluation? | The ground truth will not be made available to the performer before the evaluation. This may be part of the feedback provided to performers at the end of each evaluation. | 2/12/10 |
| 070 | Is a team including a Federally Funded Research and Development Center (FFRDC) appropriate? | No. See Section 3.A (Eligible Applicants). Federally Funded Research and Development Centers are not eligible to participate on performer teams, either as a Prime, as a sub-contractor, or in any other capacity as a member of a performer team. | 2/12/10 |
| 071 | We want to consider very scalable algorithms, possibly incorporating cloud computing. Can we experiment with cloud computing using BLACKBOOK? | Yes, BLACKBOOK can support this. | 2/12/10 |
| 072 | Is Java code and BLACKBOOK integration required just for the first cycle or as the primary delivery for all cycles? Can data storage and analysis systems separate from BLACKBOOK be proposed for this funding? | See Amendment 1, 4.B.1 Section 3, bullet E: "Software for the alignment prototype may be provided in the computer language(s) that the contractor believes to be appropriate. New software for the advanced analytic prototype must be delivered in Java (JDK 6 or higher), but it is permitted to use the Java Native Interface (JNI) ("wrappers") to utilize legacy software that may have been written in other language(s)." (See: 1.A.4.1, 1.A.5.2) BLACKBOOK integration is required for the analytic prototypes in all cycles. BLACKBOOK will be the only framework used | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|---|---|---|
| | | during the KDD evaluations. While performers can use data storage and analysis systems separate from BLACKBOOK that are already available to them, the KDD program will not fund any development of such systems whatsoever. | |
| 073 | Is it a requirement that the contract type for the prime contract flow to the subcontractors on the prime contractor's team? | It is not a requirement that the subcontractor have the same type of contract vehicle as the prime. However, be aware that the contract clauses "flow down" to the subcontractor with whatever contract vehicle they are given. | 2/12/10 |
| 074 | The BAA states that "Subcontractor proposals should include interdivisional work transfer agreements (ITWA) or similar arrangements." Please clarify the requirement for subcontractor proposals. | See Section 4.B.2/Section 2 (Detailed Estimated Cost Breakdown). The subcontractor's proposal should have the same level of details in their cost proposals that is provided by the prime contractor. This would include the direct labor rates (unburdened), indirect rates, mix and quantity of labor; any material costs, and includes backup information such as vendor quotes and catalog pages. This also applies for any interdivisional transfers. | 2/12/10 |
| 075 | Will the delivery of the prototypes for the evaluation exercise take the form of shipping the hardware with the installed prototypes to the evaluation site? Or should we alternatively budget for the purchase of our own equivalent hardware for the development phase and ship the prototype as software installations? | See 1.A.5.1 (Alignment Process and Alignment Prototype), Bullet 6, and Section 4.D (Other Submission Requirements). If the performer is using the GFE computer system, then prototypes are a software deliverable only. If a performer chooses to use a non-GFE computer system for their alignment prototype, then the hardware (with installed software) would need to be delivered to support evaluation and the costs associated with this non-GFE computer system must be part of the cost proposal. | 2/12/10 |
| 076 | Will the GFE servers provided include the classified data as well as the unclassified data? Will the servers operate at the Secret level, or Unclassified? | The GFE servers will be shipped as unclassified equipment and all data will be shipped separately for the performer to | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | | install and use. Once the classified data is loaded, the servers are considered classified. | |
| 077 | Are there existing analysis and alignment algorithms already interfaced to BLACKBOOK that could/should be leveraged for the BAA or should the contractors provide all algorithms that will be used to satisfy the requirements in the BAA? | There are no alignment algorithms interfaced to BLACKBOOK. There are currently two analysis algorithms that are interfaced with BLACKBOOK and which will be available to the performers. There is no requirement to use these analysis algorithms. The two analysis algorithms currently available in BLACKBOOK version 3.0 are: 1) WiGi - a large-scale graph viewer; and 2) Mallet - Entity extraction and disambiguation. These algorithms and documentation are available on the BLACKBOOK Wiki (see Appendix D). | 2/12/10 |
| 078 | In Section 4.B.1, Section 3, the BAA states "Government research interfaces, and planning, scheduling and control practices should be described." Could you please provide clarification on "Government research interfaces?" | Please identify the people and mechanisms that you plan to utilize in technical communication with the government during the program. | 2/12/10 |
| 079 | In Section 4.B.2 (Section 2), the BAA requires "Major program tasks by fiscal year." Is this referring to the Government fiscal year or the Contractor fiscal year? | Section 4.B.2 (Section 2: Detailed Estimated Cost Breakdown, bullet (2) Major Program Tasks by Fiscal Year) is a requirement for a high-level ("Major Program Tasks") cost estimate breakdown by Government fiscal year. | 2/12/10 |
| 080 | In reference to Section 1.A.4.1, paragraph 2,there is a statement that some of the datasets provided for research and development purposes are classified Secret. While we have facilities and personnel with TS clearances, we do not have a facility that is approved for storage of classified material. Is it feasible to implement the prototype using just the unclassified datasets? Can a facility be provided where | (See: 1.A.2, 3.D.1, 4.B.1 Detailed Proposal Information:H, 4.B.1 Security Plan, 5.A.6) KDD requires the prime contractor for each performer team to have a facility cleared at the SECRET level at the time their proposal is submitted. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | cleared personnel can install some hardware, have a workspace, and work with the classified datasets? | | |
| 081 | Section 1.C.2 of the solicitation asks for specification of waypoints in addition to milestones. In which section(s) of the proposal should the waypoints be specified? | Offerors should provide the descriptions of their waypoints in Volume 1, Technical and Management Proposal. | 2/12/10 |
| 082 | Section 1.C.2 asks for specification of waypoints between research milestones and prototype evaluation, to coincide with Technical Exchange Meetings (TEMs). Does all active research work need to have waypoints or milestones at each TEM? For example, if milestones were proposed for months 18 and 33, do we have to specify waypoints for TEMs at months 21 and 30 or could we just specify a waypoint for one of these TEMs? | (See Section 1.C.2) "Within each cycle, the KDD Program will have two project TEMs at the performer's facility. The offeror's proposed research waypoints and milestones should coincide with these reviews." | 2/12/10 |
| 083 | Will the Government accept a subcontractor's time & materials (T&M) GSA schedule rates in lieu of cost data in the prime contractor's cost proposal, where the prime anticipates the award of a time and materials contract? | A subcontractor's proposal can include its GSA schedule rates, but the proposal must also include information on the overall level of effort by the subcontractor, and an explanation of why this type of contract is being considered. | 2/12/10 |
| 084 | In Section 4.B.2 Cost Proposal, Section 2 Detailed Estimated Cost Breakdown, bullets (3) and (4), the BAA requires "an itemization of major subcontracts and equipment purchases' and "information technology purchases." Please confirm the Government desires a matrix showing the total cost by contract year for each individual subcontractor and for all contractor-furnished equipment. | Yes. Concerning material/equipment purchases, be sure to include information describing intended timelines for purchases, whether the items will be invoiced to the contract or to Internal Research & Development accounts, as well as backup documentation such as vendor quotes, catalog pages, etc. These estimates should be provided by contract year. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 085 | In reference to researchers being provided a number of data sets and questions to solve for each yearly evaluation: For each year after the first (evaluation), will the data sets change? Will there be more data sets added to the existing ones? Or, will the data sets remain static? In other words, will the data models created for Year One be applicable in Year Two? | See 1.A.4.1 (The KDD Cycle) You should assume there is a completely new problem with new data sets in each cycle, and that new models will have to be created each time. | 2/12/10 |
| 086 | In connection with Question 85, will the questions change? Will the analytical questions expand? Or will we be trying to answer the same questions with additional/different data? | See 1.A.4.1 (The KDD Cycle). You should assume there is a completely new problem with new data sets and questions in each cycle. | 2/12/10 |
| 087 | Is research in geospatial reasoning a mandatory research topic to be included in this program (i.e., is it required and within the scope of KDD) or will performers make use solely of existing technology to handle geospatial-reasoning tasks? | "Geospatial reasoning" is a broad technical area with significant existing technologies that are widely available. KDD funded research in this technical area is not precluded, but it is not a major research focus of the KDD program. | 2/12/10 |
| 088 | Will location identification/extraction from raw text be a research topic? In other words, is KDD interested about "location extraction from raw text?" | Location identification/extraction from raw text is a technical area with significant existing technologies that are widely available. KDD funded research in this technical area is not precluded, but it is not a major research focus of the KDD program. | 2/12/10 |
| 089 | With regard to geospatial queries (e.g. what significant events have occurred near this location?), can we investigate the trade-off between result quality and performance? In other words, can we use geospatial precise features (e.g., polygon, polyine) rather than bounding box/rectangle in our experiments? | This is not a research focus of the KDD program. | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| 090 | Will any output of this project carry dissemination controls, or is IARPA reserving the right to assign any dissemination controls to such output? | See 6.B.7 (Export Control), 6.B.6 (Publication Approval). Data and reports that are produced by working with KDD classified data sets will generally be classified as SECRET//NOFORN. The dissemination controls will be included in the Contract Data Requirements List (CDRL) that accompanies each awarded contract. | 2/12/10 |
| 091 | Outside of the work involving access to classified data and systems, does IARPA have any objection to the use of non-US persons to deliver against this award? | See 3.A (Eligible Applicants): "Foreign participants and/or individuals may participate to the extent that such participants comply with any necessary Non-Disclosure Agreements, Security Regulations, Export Control Laws and other governing statutes applicable under the circumstances." | 2/12/10 |
| 092 | The following text from Page 5 states "While the data model may be implicit, often it is formally defined with an explicit ontology that may or may not be readily available", is the contemporary definition of implicit used in the Semantic Web community assumed here, i.e. "implicit" means that it does not resolve to a TBox on a web server? Thus, there is always an ontology for each ingested data artifact, where "explicit" means the ontology is known publicly and "implicit" means it is not known publicly, but is present internally. | See 1.B (Data Description): "Data models can be expressed explicitly, as a formal ontology, or implicitly; If formally expressed, the model or the ontology may or may not be known." The use of the term "implicit" in the quote was common usage and not intended to convey a specific technical definition. The data sets that are the focus of the KDD program are real; with some, perhaps many, data sets we may have no idea (a priori) whether or not a formal ontology even exists. | 2/12/10 |
| 093 | Is it expected that our cost proposal and performance period are for a 51 month program? | Yes. The cost proposal is for the performance period of 51 months. | 2/12/10 |
| 094 | The GFE computer system is provided to a performer to support the performer's development and test of the performer's software using classified data sets. This requires the classified GFE system to be appropriately | No, offerors should not assume that they can move the single GFE computer system back and forth between the classified mode of use and the unclassified mode of use. Offerors may want to consider one of the following options: | 2/12/10 |

| # | Question | Answer | Date Posted |
|---|----------|--------|-------------|
| | protected, and its users must be appropriately cleared. Much of the KDD-supported research and development will be done on unclassified computer systems with unclassified data sets. It would be very convenient for the performer to be able to execute significant aspects of the algorithm research and development on the GFE computer system running as an unclassified system, using only unclassified data, and accessed by all members of the performer's team. Can the performer move the single GFE computer system back and forth between the classified mode of use and the unclassified mode of use? | 1. One additional copy of the GFE computer system may be requested, with appropriate justification (See 5.A.2: "The requirement for and the anticipated use or integration of Government Furnished Property (GFP) including all equipment, facilities, information, etc., is fully described including dates when such GFP, GFE (Government Furnished Equipment), GFI (Government Furnished Information) or other similar Government-provided resources will be required."); 2. The offeror may propose to purchase an additional copy themselves. Offerors will need to include a cost estimate in their cost proposal; See Section 4.B.2. The complete specification (hardware and software) of the GFE computer system will be made available at kickoff. The offeror may assume that a single such system costs approximately $8000. | |
| 095 | Does a Top Secret Facilities Security Clearance satisfy the requirement for a SECRET level facility? | Yes. | 2/12/10 |
| 096 | Our understanding is that the Defense Security Service (DSS), Office of the Designated Approving Authority (ODAA) will accredit information systems at the SECRET level only after a DD254 is provided for the contract. Are offerors required to provide documentation that confirms SECRET level accreditation of proposed information systems at time of proposal submission? | Those offerors wishing to use an information system other than the GFE information system for classified work, must provide certification that the information system that they intend to use is approved to operate at the SECRET level at the time that they submit their proposal. Those offerors electing to utilize the GFE information system will not require an information system pre-accreditation, and will be allowed sufficient time and assistance to acquire the required SECRET level Information System accreditation after receipt of the GFE equipment. | 2/12/10 |